1Ш

УДК 004.048

Р.Г. Шыхалиев
Институт информационных технологий НАН Азербайджана, г. Баку
ramiz@science.az

О применении интеллектуальных технологий в мониторинге компьютерных сетей

В статье рассматриваются некоторые вопросы применения интеллектуальных технологий в сетевом мониторинге компьютерных сетей (КС), которые являются очень важными дополнениями к системам сетевого мониторинга. А именно вопросы использования мобильных агентов в системах сетевого мониторинга и управления КС, а также методов машинного обучения для классификации сетевого трафика КС.

Введение

Сегодня все больше и больше компьютерным сетям (КС) доверяют решение многих задач жизнедеятельности, таких, как, управление энергетической системой, технологическим процессом, личная и корпоративная коммуникация, банковское дело, торговля, посещение магазинов и т.д. При этом, с каждым днем масштаб и сложность современных КС растут, а также увеличивается скорость передачи данных в них. Кроме того, изменяются характер и объем сетевого трафика, а также создаются новые технологии построения КС (например, оптические, беспроводные и т.д.), что приводит к расширению диапазона неисправностей, происходящих в КС, растет размер ущерба из-за их простоев. В результате этого управление КС становится очень трудоемкой задачей и требует больших человеческих ресурсов. При таких условиях выполнение основных обязанностей системных администраторов КС зависит от результатов сетевого мониторинга.

Цель работы – проанализировать специфические особенности задач сетевого мониторинга КС, а именно рассмотреть вопросы использования агентов для централизации управления мониторинга КС, а также методы машинного обучения для классификации сетевого трафика КС.

Сетевой мониторинг – это слежение и анализ состояний и поведения управляемых КС. Сетевой мониторинг является источником получения объективных данных о состоянии и функционировании КС, без чего трудно принять обоснованные решения по их управлению. Кроме того, без проведения сетевого мониторинга трудно сделать объективное заключение о конфигурации сетевого аппаратного и программного обеспечения КС, а также о результатах проведенных в них изменений и обеспечить безопасность КС. Поэтому состояние сегодняшних КС должно постоянно контролироваться, что невозможно без сетевого мониторинга. Однако постоянный мониторинг КС не всегда выгоден и не нужен, так как может привести к большим издержкам из-за огромности собираемого сетевого трафика, объема передаваемых по сети мониторинговых данных и т.д.

Анализ существующих методов мониторинга КС показал, что они имеют ряд недостатков. Один из недостатков заключается в том, что эти подходы имеют характер централизации, при которой все вычислительные нагрузки сосредоточиваются на одном компьютере. В результате огромный объем мониторинговых данных должен быть

передан в центр мониторинга для дальнейшей обработки, что загружает центр мониторинга, а также каналы самой КС. Другим недостатком является то, что существующие подходы сетевого мониторинга собирают и отображают сетевой трафик и данные без использования эффективных методов добычи данных (Data mining). Третьим недостатком является отсутствие возможности расширения функций мониторинга при необходимости и т.д.

Кроме того, растущая сложность и масштаб сегодняшних КС, а также количество используемых в них сетевых сервисов и приложений, механизмов защиты, аппаратного и программного обеспечения могут привести к тому, что сам процесс мониторинга будет генерировать большой трафик и потреблять огромные системные ресурсы. Например, имеющиеся в КС серверы, межсетевые экраны, маршрутизаторы, системы обнаружения вторжения и т.д. генерируют непрерывный поток мониторинговых сообщений. В конечном счете это может привести к нарушению нормальной работы КС, поэтому необходимо создание новых технологий мониторинга, которые позволят проводить децентрализованный мониторинг КС.

Поэтому для повышения эффективности мониторинга КС актуальным является использование интеллектуальных технологий (ИТ), так как они способны значительно упростить и облегчить процесс мониторинга КС. Кроме того, использование ИТ позволяет минимизировать роль человека при мониторинге КС, уменьшить потери нужной информации, минимизировать влияние мониторинговой системы на нормальную работу КС и т.д.

1 Управление и мониторинг КС

Существуют различные функциональные требования к системам управления КС. Самая известная классификация функций систем управления КС определена ISO (The International Organization for Standardization, International Standards Organization), которая состоит из управления конфигурацией, неисправностями, производительностью, безопасностью и учетом сетевых ресурсов [1], [2].

Традиционные системы управления КС основываются на SNMP (Simple Network Management Protocol) IETF или CMIP (Common Management Information Protocol) OSI [3] и состоят из двух компонентов: менеджера и агентов. В обоих подходах управление КС осуществляется централизованным способом. На рис. 1 изображена схема типичной системы управления сетью.

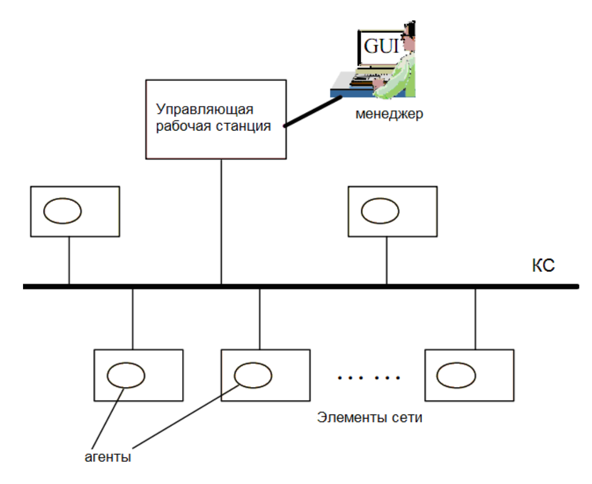


Рисунок 1 – Компоненты типичной системы управления КС

Приложения, находящиеся в управляющей рабочей станции, играют роль менеджера и позволяют администратору КС с помощью GUI (Graphical User Interface) осуществлять мониторинг КС. Агенты – это часть процессов сервера, выполняемых на объектах управляемой КС. Агенты собирают данные о сетевых устройствах и хранят их в MIB (Management Information Base) базах, которые поддерживают SNMP или CMIP-протоколы [4], [5].

Главной целью систем управления КС является гарантирование качество услуг (то есть QoS), которое предоставляет КС. Для этого менеджеры КС должны провести мониторинг и управлять взаимосвязанными элементами КС. Сетевой мониторинг проводит наблюдение и анализ состояния и поведения сетевых устройств, которые входят в конфигурацию КС и должны ею управляться. Поэтому точный и эффективный мониторинг является очень важным для выполнения различных функций управления КС.

Современные системы мониторинга/управления КС имеют централизованную структуру, в которой основная часть функций мониторинга, интеллектуализации и обработки осуществляется с помощью приложений управляющих рабочих станций. Поэтому мониторинг всех соединений и обработка собранных мониторинговых данных осуществятся на управляющей рабочей станции, что может привести к эффекту «бутылочной горалы» и единственной точки отказа. Такая централизация снижает эффективность сетевого мониторинга, особенно в высокоскоростных КС, а также не пригодна для больших и сложных КС.

Так как приложения менеджера могут взаимодействовать с элементами КС только через низкоуровневые интерфейсы общего назначения, то решение даже незначительной задачи мониторинга потребует передачи большого объема «сырых» SNMP-данных в управляющую рабочую станцию, что может привести к непроизводительным издержкам и существенному снижению производительности управляемой КС.

Управление КС можно разделить на два логических компонента, то есть на мониторинг сети и контроль сети. Фактически каждая из пяти упомянутых выше функциональных областей может включить в себя эти компоненты. В свою очередь функции сетевого мониторинга КС можно разделить на два этапа: обследование (измерение) и анализ. А к основным задачам сетевого мониторинга КС относятся: 1) мониторинг безопасности КС, который заключается в обследовании и анализе доступа пользователей в КС, чтобы обнаружить ошибочные, незаконные или злонамеренные пользовательские операции, которые могут поставить под угрозу безопасность КС; 2) мониторинг производительности КС, который заключается в обследовании и анализе индикаторов, имеющих отношение к производительности КС, таких, как утилизация, пропускная способность, доступность и т.д., чтобы обнаружить ухудшение производительности каналов и сетевых устройств КС; 3) мониторинг неисправностей в КС, который заключается в обследовании и анализе индикаторов (известных так же, как симптомы), имеющих отношение к неисправностям, чтобы обнаружить потенциальные неисправности; и т.д.

Контроль сети способен изменить или переформировать определенные части КС, чтобы восстановить работу сети при обнаружении некоторых отклонений от нормы и сообщить об этом сетевому монитору.

2 Интеллектуальный сетевой мониторинг КС с помощью мобильных агентов

Основные проблемы мониторинга распределенных и неоднородных КС связаны с тем, что топология КС может постоянно изменяться, полоса пропускания – ограниченной, латентность – высокой, коммуникационные каналы – недоступными и т.д. Ис-

пользование мобильных агентов (МА) в системах сетевого мониторинга КС позволит решить эти проблемы, то есть заранее получить информацию об этих проблемах и реагировать на них.

МА – это мультиагентные системы, которые включают в себя все признаки агента [6]. Эти агенты являются интеллектуальными, могут содержать код управления и переходить от узла к узлу КС, пока не выполнят свою задачу, и сохраняют информацию об их состоянии в отдельности. Для достижения своих целей они также общаются с другими агентами.

МА функционируют следующим образом. Изначально МА создаются и хранятся на отдельном компьютере, так называемой исходной машине (home machine). Далее агент отправляется на удаленный компьютер, так называемый хост МА, для выполнения. Хост МА также называют сервером МА. Вместе с агентом хосту отправляют весь код МА и информацию о состоянии МА. Хост предоставляет агенту подходящую среду для выполнения. При этом для выполнения своей задачи МА используют ресурсы хоста (ресурсы центрального процессора, память и т.д.). После завершения своей задачи на хосте МА переходят на другой компьютер. При этом миграция МА продолжается до тех пор, пока МА не возвратится к исходной машине после завершения выполнения задачи на последней машине в маршруте. При этом жизненный цикл МА состоит из следующих этапов (рис. 2):

1. МА создается на исходной машине (home machine).
2. МА посылается на хост МА для выполнения.
3. Агент выполняется на хосте А.
4. После выполнения агент создает свою копию, затем одна копия отправляется хосту В, а другая – хосту С.
5. Создание копий осуществляется на соответствующих хостах.
6. После выполнения на хостах В и С МА отправляются исходной машине (home machine).
7. Исходная машина (home machine) принимает агенты и анализирует собранные ими данные.

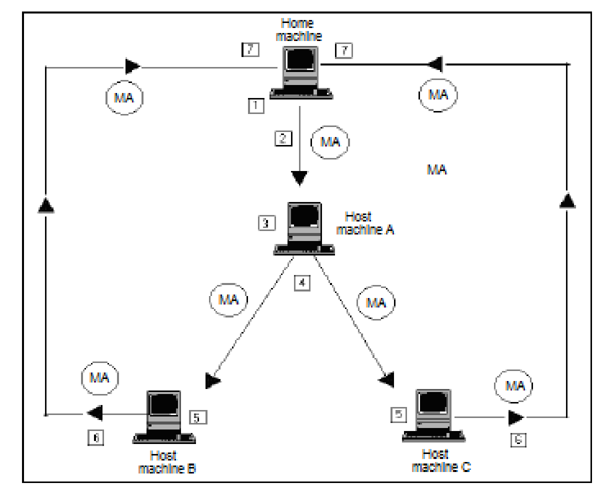


Рисунок 2 – Жизненный цикл МА

МА имеют некоторые преимущества, связанные со снижением загрузки и латентности сети, инкапсуляцией протоколов, асинхронным и автономным выполнением и отказоустойчивостью.

Наряду с преимуществами у МА имеются недостатки. Основные недостатки связаны с наличием рисков безопасности использования МА и безопасности мобильной

среды вычисления, то есть злонамеренная МА может повредить хост. Например, вирус может замаскироваться под МА и распространяться в КС и нарушить работу хостов. С другой стороны, злонамеренные хосты могут нарушить функции МА.

Существуют различные механизмы защиты от злонамеренных действий. Одна из предложенных схем защиты от злонамеренных действий заключается во введении механизма трассировки действий МА на каждом хосте. Используя трассировку, можно обнаружить как злонамеренные действия МА, так и злонамеренные хосты. Однако, несмотря на существенное развитие криптографии, существуют проблемы, связанные с безопасностью МА [7], [8].

Имеются различные подходы к управлению КС с использованием МА. Первый подход основан на создании архитектуры, которая включает менеджеров, серверы и управляющие агенты [9]. В этой архитектуре нагрузка управления сетью одинаково распределена между менеджерами и серверами. Серверы могут тиражировать базы данных. Для передачи, получения и хранения МА применяется структура, использующая среду MAGENTA (Mobile Agent for Administration).

Второй подход заключается в создании платформы для создания и управления МА [10]. Эта платформа соответствует стандарту MASIF (Mobile Agent System Interoperability Facility) [11]. Были созданы приложения для управления основными функциями агентов. Эти агенты выполняют простые задачи, но их комбинация может быть использована для выполнения сложных задач управления.

В третьем подходе предлагается MABNM (Mobile Agent Based Network Management) – структура, которая поддерживает как инфраструктуры мобильных кодов, так и средств имитации и управления КС [12]. MCD (Mobile Code Daemon) запускает JVM (Java Virtual Machine) на каждом управляемом элементе КС и для получения Java-ответов слушает UDP или TCP-порты, на которые отправлены запросы.

В работе [13] для сетевого мониторинга предлагается структура, состоящая из четырех компонентов: приложение менеджера (Manager), сервер MAS (Mobile Agent Server), генератор MAG (Mobile Agent Generator) и МА (Mobile Agents). Приложения менеджера управляют мониторингом элементов сети и имеют GUI. MAS получает МА, выполняет их коды, присваивает им значение и отправляет их. А MAG создает МА, соответствующие требованиям определенных сервисов.

3 Классификации сетевого трафика КС с использованием методов машинного обучения

В основном сетевой трафик в КС состоит из трафиков клиентов, серверов и приложений и характеризуется множеством признаков, которые используются для сетевого мониторинга КС. В качестве таких признаков могут применяться DNS-запросы, DHCP-запросы, DHCP-ответы, WINS-трафики, числа пакетов, объем и скорость входящего и выходящего трафиков, IP-адрес отправителя и получателя, MAC-адрес хостов, виды используемых протоколов (например, HTTP, FTP, SMTP и т.п.) и приложений, время и т.д.

Классификация сетевого трафика в реальном масштабе времени является одной из основных проблем сетевого мониторинга КС, который позволит решить проблемы эффективного управления КС.

Классификация сетевого трафика позволяет идентифицировать используемые приложения на основании того, что в основном приложении используют конкретные «известные» порты TCP или UDP (обычно информация о номере порта имеется в заголов-

ках TCP или UDP-пакетов). Однако все номера портов, используемые большинством приложений, предсказать невозможно [14]. Поэтому нужны более эффективные методы классификации сетевого трафика, которые позволяют определить тип приложения на основе данных, имеющихся в основной части TCP или UDP-пакетов (или на основе известных поведений протоколов) [15]. Однако из-за детальности проверки содержимого пакетов уменьшается эффективность таких методов классификации сетевого трафика.

Большинство исследователей считают методы машинного обучения (МО), которые являются частью дисциплины искусственного интеллекта, более подходящими для классификации сетевого трафика. В 1990 году в работе [16] был предложен инспектор сетевого трафика, основанный на методах МО и предназначавшийся для минимизации длительности вызовов в телекоммуникационных сетях с канальной коммутацией. Эта работа является началом применения методов МО в области телекоммуникационных сетей. В 1994 году МО было использовано для классификации интернет-трафика в целях обнаружения вторжения [17]. Эта работа положила начало применению методов МО для классификации интернет-трафика.

В общем МО – это процесс поиска и описания шаблонов в структуре заданной выборки наборов данных. При этом на вход МО подаются некоторые образцы в виде набора данных (то есть примеры). Каждый образец характеризуется значениями свойств (то есть атрибутами), которые выражают различные аспекты образца. Например, в КС последовательность пакетов некоторого сетевого трафика может состоять из образцов, при этом атрибутами этих образцов может быть среднее время между пакетами или среднеквадратическое отклонение длины пакетов и т.д. Наборы данных могут быть представлены в виде матрицы взаимосвязи элементов и атрибутов [18].

Выход МО зависит от описания знаний, которым должна быть обучена модель МО. А представление результата процесса обучения (синтаксис и семантика) в значительной степени зависит от используемого подхода МО.

Обучение с учителем создает структуры знаний, которые используются для отнесения новых образцов в заранее определенные классы [19]. Обучение машины проводится с представлением к ее входу наборов типовых примеров, которые относятся заранее к определенному классу. Результатом процесса обучения является построение модели классификации на основе анализа и обобщения представленных образцов.

Формально классификация определяется следующим образом. Пусть T – набор данных примера, который является множеством пар входов/выходов классификатора, то есть $T = \{< x_1, y_1 >, < x_2, y_2 >, <>, \ldots, < x_n, y_m >\}$, где x_i – вектор входных атрибутов, соответствующий i-му образцу, и y_i – вектор значения выходных классов. Задача классификации состоит в следующем. В наборе данных примера T требуется найти такую функцию от входных атрибутов $f(x)$, которая точно определяет соответствующий выходной класс y для всех новых значений x. Причем y принимает дискретные значения из множества $\{y_1, y_2, \ldots, y_m\}$, которое состоит из всех заранее определенных значений классов.

На рис. 3 показана модель обучения и тестирования классификатора. В этом примере показан классификатор, который распознает трафики некоторых классов приложений (например, трафики онлайновых игр в реальном масштабе времени) в общем сетевом трафике. При этом для оптимального обучения и тестирования классификатора МО с учителем, классификатору должны быть представлены ранее классифицированные примеры трафиков двух типов. Первый тип – это трафики приложений, относящихся

к представляющему интерес классу, которые должны быть идентифицированы (например, трафики онлайновых игр), а второй тип – это примеры трафиков других приложений, которые могут быть идентифицированы в будущем.

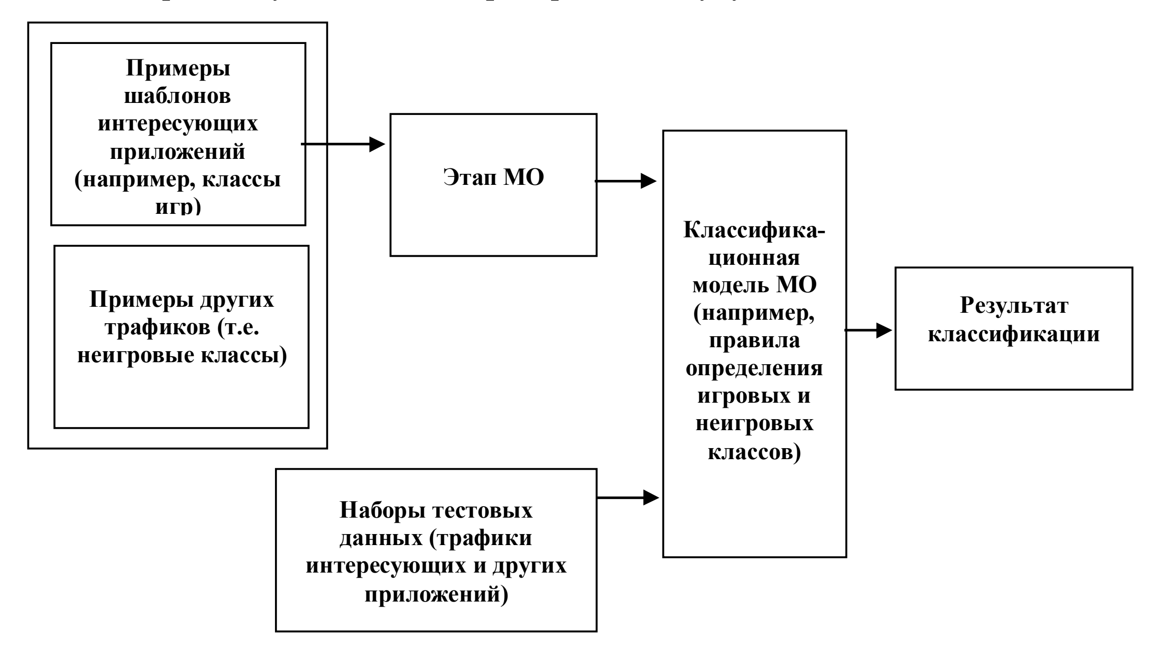


Рисунок 3 – Модель обучения и тестирования классификатора МО с учителем

Процесс обучения классификатора МО с учителем заключается в следующем (рис. 4). Сначала на этапе «запись трафика» записывается полный сетевой трафик, который содержит и трафики приложений, относящихся к интересующему нас классу (например, трафики онлайновых игр), и трафики других приложений (такие, как HTTP, DNS, SSH и/или peer2peer и т.д.).

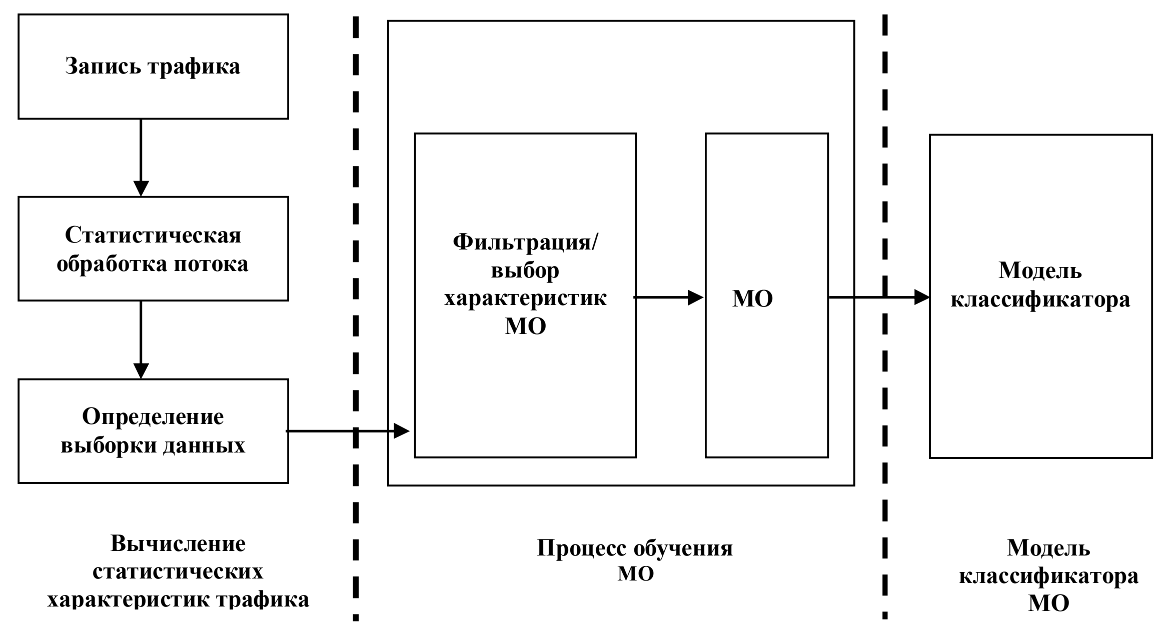


Рисунок 4 – Процесс обучения классификатора МО с учителем

На шаге «Статистическая обработка потока» вычисляются статистические свойства потока (такие, как средний размер пакетов, стандартное время задержки между пакета-

ми, полный размер трафика и т.д.), которые используются для генерации шаблонов трафика. Далее для снижения области поиска для алгоритма обучения МО, то есть снижения размерности выборки обучающих данных, на шаге «Выборка данных» выделяются те признаки, которые представляют интерес. На этом шаге выделяются статистические характеристики подмножества образцов, относящихся к различным классам приложений, которые используются классификатором в процессе обучения. А на выходе работает модель классификации.

Существует ряд алгоритмов классификации на основе обучения с учителем, которые отличаются, главным образом, используемыми классификационными моделями и алгоритмами поиска решений. В качестве примера можно привести дерево решений с учителем и простые алгоритмы классификации Байеса [18], [20].

Заключение

В работе были проанализированы особенности задач сетевого мониторинга КС, и можно сделать вывод о том, что задачи, решаемые в сфере управления/мониторинга КС, сложны и разнообразны. Эффективное их решение невозможно без применения интеллектуальных технологий.

Основные проблемы сетевого мониторинга КС связаны с тем, что топология КС может постоянно меняться, полоса пропускания – быть ограниченной, латентность – высокой, коммуникационные каналы – недоступными и т.д. Использование МА в системах сетевого мониторинга КС позволит решить эти проблемы, то есть заранее получить информацию об этих проблемах и реагировать на них, а также обеспечить децентрализованное управление/мониторинг КС.

Классификация сетевого трафика в реальном масштабе является актуальной задачей эффективного сетевого мониторинга КС. Она позволяет идентифицировать используемые приложения на основании «известных» номеров портов TCP или UDP (обычно информация о номере порта имеется в заголовках TCP или UDP-пакетов), используемых ими. Однако все номера портов, используемые большинством приложений, предсказать невозможно. Поэтому нужны более эффективные методы классификации сетевого трафика, которые позволяют определить тип приложения на основе данных, имеющихся в основной части TCP или UDP-пакетов (или на основе известных поведений протоколов).

Большинство исследователей рассматривают методы МО как более подходящие для классификации сетевого трафика. Однако каждый алгоритм МО имеет различные подходы к классификации и назначению приоритетов, наборам особенностей, которые приводят к различным способам обучения и классификации.

Результаты исследований, проведенных в области классификации сетевого трафика на основе МО, могут быть использованы при обнаружении вторжений в КС, аномалий в поведении пользователей, то есть в определении профилей пользователей, а также в создании профиля КС.

Литература

1. Stallings W. SNMP, SNMPv2 and CMIP: the practical guide to network management standards / W. Stallings. – Addison-Wesley, Reading, Mass., 1993.
2. Mamdani E.H. The Management of Telecommunication Networks / E.H. Mamdani, R. Smith and J. Callaghan. – Ellis Horwood Limited, 1993.
3. Warrier U. The Common Management Information Services and Protocols over TCP/IP (CMOT) RFC 1 095 / U. Warrier, L. Besaw.

4. Yemini Y. The OSI network management model / Y. Yemini // IEEE communication. – May 1993. – P. 20-28.
5. Vuanh O. Mobile software agents: An Overview / O. Vuanh and K. Ahmed // IEEE Communication magazine. – July 1998. – P. 25-37.
6. Makki S. Wunnava, Application of Mobile Agents in Managing the Traffic in the Network and Improving the Reliability and Quality of Service, IAENG / S. Makki, V. Subbarao // International Journal of Computer Science, 32:4, IJCS_32_4_16
7. Hohl F. A Model of Attacks of Malicious Hosts Against Mobile Agents / F. Hohl // Proceedings of the ECOOP Workshop on Distributed Object Security and 4th Workshop on Mobile Object Systems : Secure Internet Mobile Computations. – INRIA, France, 1998. – P. 105-120.
8. Hohl F. Time Limited Blackbox Security : Protecting Mobile Agents From Malicious Hosts / F. Hohl // Mobile Agents and Security, LNCS 1419. – Springer-Verlag, 1998. – P. 92-113.
9. Sahai A. Towards Distributed and Dynamic Network Management / A. Sahai, C. Morin // Proceedings of the 1998 IEEE Network Operations and Management Symposium. – New Orleans, USA, February 1998. – Vol. 2. – P. 455-464.
10. Puliafito A. Using Mobile Agents to Implement Flexible Network Management Strategies / A. Puliafito and O. Tomarchio // Computer Communication Journal. – April 2000. – Vol. 23(8). – P. 708-719.
11. MASIF: The OMG Mobile Agent System Interoperability Facility / D. Milojicic, M. Breugst, I. Busse [et al.] // Proceedings of the Second International Workshop on Mobile Agents. – Stuttgart, Germany, 1998. – P. 50-67.
12. Kona M.K. A Framework for Network Management Using Mobile Agents / M.K. Kona, C-Z. Xu // Proceedings of the First IEEE Int'l Workshop on Internet Computing and E-Commerce. – San Francisco, USA, April, 2001.
13. Advanced Network Monitoring Applications Based on Mobile / D. Gavalas, D. Greenwood, M. Ghanbari, M. O'Mabony // Intelligent Agent Technology, Computer Communications Journal. – April 2000. – Vol. 23, № 8. – P. 720-730.
14. Is P2P dying or just hiding? / T. Karagiannis, A. Broido, N. Brownlee and K. Claffy // Proceedings of the 47th annual IEEE Global Telecommunications Conference (Globecom 2004). – Dallas, Texas, USA, November/December 2004.
15. Sen S. Accurate, scalable in network identification of P2P traffic using application signatures, in WWW 2004 / S. Sen, O. Spatscheck, and D. Wang. – New York, NY, USA, May 2004.
16. Silver B. Netman: A learning network traffic controller / B. Silver // Proceedings of the Third International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems, Association for Computing Machinery. – 1990.
17. Frank J. Machine learning and intrusion detection : Current and future directions / J. Frank // Proceedings of the National 17th Computer Security Conference. – Washington, D.C., October 1994.
18. Witten I. Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations (Second Edition) / I. Witten and E. Frank. – Morgan Kaufmann Publishers, 2005.
19. Reich Y. The formation and use of abstract concepts in design / Y. Reich and J.S. Fenves // Fisher D.H. Concept Formation : Knowledge and Experience in Unsupervised Learning / D.H. Fisher, M.J. Pazzani. – Morgan Kaufmann, 1991.
20. Fisher H.D. Concept Formation : Knowledge and Experience in Unsupervised Learning / H.D. Fisher, J.M. Pazzani, P. Langley. – Morgan Kaufmann, 1991.

R.G. Shikhaliev

Application of Intelligent Technologies at the Network Monitoring of Computer Networks

In this article some issues of application of intelligent technologies at the network monitoring of computer networks (CN) are discussed, which are very important additions to the network monitoring systems. Namely, using of mobile agents in the network monitoring systems and management of CN, as well as methods of machine learning for classification of CN network traffic.

Статья поступила в редакцию 28.09.2010.